



**Federal Aviation
Administration**

DOT/FAA/AM-11/18
Office of Aerospace Medicine
Washington, DC 20591

Evaluation of a Fatigue Countermeasures Training Program for Flight Attendants

Erica L. Hauck
Katrina Bedell Avers
Joy O. Banks
Lauren V. Blackwell

Civil Aerospace Medical Institute
Federal Aviation Administration
Oklahoma City, OK 73125

November 2011

Final Report

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications Web site:
www.faa.gov/library/reports/medical/oamtechreports

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-11/18	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Evaluation of a Fatigue Countermeasures Training Program for Flight Attendants		5. Report Date November 2011	6. Performing Organization Code
7. Author(s) Hauck EL, Avers KB, Banks JO, Blackwell LV		8. Performing Organization Report No.	
9. Performing Organization Name and Address FAA Civil Aerospace Medical Institute P.O. Box 25082 Oklahoma City, OK 73125		10. Work Unit No. (TRAIS)	11. Contract or Grant No.
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591		13. Type of Report and Period Covered	14. Sponsoring Agency Code
15. Supplemental Notes Work was accomplished under approved task AM-A-08-HRR-521			
16. Abstract <p>There is a growing population of employees that work non-traditional hours in around-the-clock operations. Cabin crew/flight attendants are part of this population and work highly variable schedules that include extended duty days, time zone changes, night schedules, and on-demand calls. These schedules conflict with the body's natural mechanisms for managing sleep and alertness, and often result in fatigue. Fatigue countermeasure training may be necessary for the health, well-being, and safety of workers. The current research evaluated a comprehensive fatigue countermeasure training program for flight attendants using a theoretically grounded taxonomy of training criteria. Alternative evaluation strategies were also utilized to improve traditional pretest-posttest designs and provide convergent evidence of training effectiveness.</p>			
17. Key Words Training, Flight Attendants, Cabin Crew, Education, Fatigue, Countermeasures		18. Distribution Statement Document is available to the public through the Defense Technical Information Center, Ft. Belvoir, VA 22060; and the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No. of Pages 14	22. Price

Form DOT F 1700.7 (8-72)

Reproduction of completed page authorized

ACKNOWLEDGMENTS

The development of this report was supported by the Human Factors Research Division of the Civil Aerospace Medical Institute and the ATO Human Factors Research and Engineering Group

CONTENTS

INTRODUCTION	1
Fatigue Countermeasures Training Evaluation.....	1
METHOD.....	2
Course Development.....	2
Training Delivery	3
Training Evaluation.....	3
RESULTS	4
Hypothesis Tests	4
DISCUSSION.....	7
REFERENCES.....	8

EVALUATION OF A FATIGUE COUNTERMEASURES TRAINING PROGRAM FOR FLIGHT ATTENDANTS

INTRODUCTION

Cabin crew/flight attendants perform a number of pre-flight, during flight, and post-flight checks to ensure passenger safety. They work highly variable, and often more extreme schedules than pilots and are sensitive to extended schedules, time zone changes, night schedules, and on-demand calls. In 2005, Congress directed FAA's Civil Aerospace Medical Institute (CAMI) to investigate fatigue in cabin crew operations. CAMI teamed with the National Aeronautics and Space Administration (NASA) Ames Research Center's Fatigue Countermeasure Group to conduct a preliminary study of fatigue and found that "flight attendant fatigue appears to be a salient issue warranting further evaluation." (Nesthus, Schroeder, Connors, Rentmeister-Bryant, & DeRoshia, 2007; p.21) The findings of that study lead to a series of congressionally mandated follow-on studies in 2008, including: a survey of field operations, a field study on the effects of fatigue, validation of models for assessing fatigue, a focused study of incident reports, a review of international policies and practices, and a review of the benefits of training for fatigue risk management. During a field study, objectively measuring sleep with actigraphy, flight attendants were found to be sleeping an average of only 5.7 hours per night on work days (Roma, Mallis, Hursh, Mead, & Nesthus, 2010). Flight attendants often worked in a fatigued state. These fatigue levels are influenced by type of operation, duty duration, continuous-duty overnights, reserve practices, reduced rest, lack of breaks, restricted rest periods, and duty report times (Avers et al., 2009b; Roma et al., 2010). Some of the key conclusions resulting from these studies indicated the industry needs to: 1) identify ways to improve schedules from a science-based approach to maximize alertness and minimize fatigue while meeting operational and economic constraints of the industry; 2) develop an adaptive fatigue mitigation safety system such as a fatigue risk management system (FRMS) that combines scientific principles and knowledge with operational support and constraints; 3) apply scientific modeling tools to maximize alertness and minimize fatigue while meeting operational and economic constraints; 4) develop and administer a comprehensive, science-based fatigue countermeasure training program; and 5) establish a flight attendant fatigue workgroup of subject matter experts, aviation stakeholders, medical and

research scientists, and aviation safety management system (SMS) experts to evaluate 14 CFR sections 121.467 and 135.273 for possible revision (Avers, Hauck, Blackwell, & Nesthus, 2009a; Avers et al., 2009b; Banks, Avers, Nesthus, & Hauck, 2009; Holcomb et al., 2009; Nesthus et al., 2007; Roma et al., 2010).

In response to one of the recommendations from these studies, this paper will evaluate the benefits of a fatigue countermeasures training program developed specifically for flight attendants.

Fatigue Countermeasures Training Evaluation

The fatigue countermeasure training evaluation was designed to establish a standard for fatigue countermeasures training programs with regard to content development and applicability across occupations that utilize non-traditional work shift schedules. A theoretically grounded taxonomy of training criteria was used to assess training success across multiple domains. This was done to improve the training process and provide a more complete evaluation of learning. The following hypotheses were developed from this discussion:

- H1: Performance on cognitive measures will improve from pretest to posttest and follow-up.
- H2: Motivation, attitude strength, and self-efficacy will improve from pretest to posttest and follow-up.
- H3: Use of fatigue countermeasures will improve from pretest to follow-up.

As discussed previously, shiftworkers are especially prone to experiencing fatigue, sleepiness, physical symptoms, and work-family conflict. This is in large part due to the mismatch of their schedules with the body's circadian rhythms and most diurnal schedules of the working/domestic/social environments; however, the degree to which these outcomes are directly affected by fatigue are relatively unknown, though improved fatigue management should minimize negative outcomes. The following hypothesis was developed from this discussion:

- H4: Fatigue, sleepiness, the experience of physical symptoms, and work-family conflict will decrease from pretest to follow-up.

Additionally, previous training evaluation methodologies were expanded and enhanced to rule out lingering threats to validity and improve confidence in

the conclusions drawn from our training evaluation. By using multiple evaluation strategies, training outcomes can be compared to provide convergent evidence for the effectiveness of the training. Two training evaluation strategies that allowed evaluators to have greater confidence in the inferences drawn from evaluation results were the *internal referencing strategy* (IRS) and the *rolling group design* (RGD).

Internal referencing is a useful variant of our pretest—posttest design in which the training evaluator includes both training-relevant and training-irrelevant items in the pretest and posttest (Haccoun & Hamiaux, 1994). The concept is that our training-relevant items should demonstrate greater improvement following the training session than the training-irrelevant items because that information is actually part of the training content. The irrelevant items serve as a proxy control group for the relevant items. Ideally, all items would be derived from the same topic area, but the information concerning training-irrelevant items would not be covered during the training course. This design is especially useful for ruling out threats to validity such as history, maturation, and testing effects and is not subject to the validity threats that typically plague between-subjects designs (Frese, Beimel, & Schoenborn, 2003; Haccoun et al., 1994). Haccoun and Hamiaux empirically tested and supported results suggesting that IRS produces similar inferences about the effectiveness of training as a pretest—posttest with control group design. However, only a handful of other published studies have used this method of evaluation (Aguinis & Branstetter, 2007; Cigularov, Chen, Thurber, & Stallones, 2008; Frese et al., 2003; Oostrom & Van Mierlo, 2008) and has sometimes been referred to as using a nonequivalent, dependent variable design (Shadish, Cook, & Campbell, 2002). The following hypotheses were developed from this discussion:

- H5: Change in performance from pretest to posttest and follow-up on declarative knowledge measures will be greater for relevant items than for non-relevant items.

The RGD is another variation of the traditional pretest—posttest design in which a group of individuals, who will eventually be trained, serve as a control group until they receive the training (Quinones & Tonidandel, 2003). If the first group to receive training is the training group and the second group to receive training is the “control group,” the design allows an evaluation of significant mean differences between: 1) the pretest—posttest performance of the training group, 2) the posttest performance of training group and the pretest performance of the control group, and 3) the pretest—posttest performance of control group (Cigularov et al., 2008). Additionally, there should be no significant difference between the

pretest scores for the training group and for the control group. This design is similar to a pretest—posttest with a nonequivalent control group and is particularly useful when the training will be repeated, with no access to a pre-designated control group. To date, Cigularov and colleagues have published the only known example of RGD. The following hypotheses were developed from this discussion:

- H6: Performance on cognitive measures will improve from pretest to posttest in the training group.
- H7: Performance on cognitive measures will improve from the pretest of the “control group” to the posttest of the training group.
- H8: Performance on cognitive measures will improve from pretest to posttest in the “control group.”

The current study demonstrates the usefulness of fatigue countermeasures training for workers with non-traditional work schedules. Specifically, our study incorporated volunteer flight attendants during the training evaluation portion of this research. Content analysis of existing fatigue-related training programs was conducted and supplemented with additional materials specific to the flight attendant workforce. Development of our training program was followed by an evaluation using a pretest—posttest follow-up training design that included internal referencing and the rolling group design recommended to protect against threats to validity. Kraiger, Ford, and Salas's (1993) taxonomy of learning outcomes was also used to thoroughly evaluate the training program.

METHOD

Course Development

A multi-method approach was used to develop recommendations for topics that should be included in a comprehensive fatigue management training program. The process began with identification of existing fatigue training programs. These were content-analyzed and used to create a basic outline for a fatigue management training program. An extensive literature review was used to supplement the basic outline with flight attendant-specific information and other, less frequently cited fatigue topics. Course content was developed using existing training programs, empirical literature, expert input, and other relevant sources. A final course content check was completed by two subject matter experts who were instructed to examine all of the training content for deficiencies, excesses, and inaccuracies.

Once the training outline had been developed, each topic area was populated with current information and research. Existing training programs that were part of the public domain, empirical literature, and experts were all

consulted to create the most current training material possible. After the information had been compiled, synthesized, and organized in a meaningful way, the entire document was reviewed by multiple experts in the field of sleep and fatigue research. Modifications were made based on this ad hoc group's feedback, and a final review was conducted. Handout materials were also created to summarize important topic areas and provide take-away information.

Training Delivery

Participants. A total of 50 domestically-based flight attendants volunteered to attend a one-day training event. To recruit participants, correspondence was sent to airlines, union representatives, and professional contacts in the aviation industry, providing information along with a Web site link to register for the training. Participants were responsible for signing up via the Web site and selecting one of the three training sessions to attend. They were provided confirmation, travel and lodging information, and a detailed itinerary via e-mail, though no monetary compensation.

Ten flight attendants participated in the first training event, 23 participated in the second, and 17 participated in the third. The mean age of our participants was 46.66 years with 72% (n=36) being female and 28% (n=14) being male. The length of time participants worked in the flight attendant field ranged from 2.83 years to 38.83 years ($M = 11.12$). Of the 50 flight attendants who participated, two were dropped from further analyses due to extensive knowledge of fatigue prior to the training (in both instances, the flight attendants represented dual roles in their organizations, contributing to a broader existing knowledge of fatigue).

Instructional Mode. A traditional PowerPoint lecture and discussion delivery method was used for the training program, with the addition of supplemental materials when appropriate (e.g., short video clips, accident reports, interactive personal experience, etc.).

Procedure. Flight attendants participated in the fatigue countermeasures training as a part of a one-day event hosted by the FAA. Prior to arrival, flight attendants were asked to complete an online survey that included questions and the various training-relevant and irrelevant pretest measures. The training lasted approximately three hours and was followed by administration of posttest measures. All participants were provided with a handout of the training materials and tools to aid fatigue prevention and management. Approximately six weeks after the initial training, participants were contacted via e-mail and asked to complete a follow-up survey. Up to two reminder e-mails were sent to encourage completion of the follow-up survey.

Training Evaluation

Training criteria were developed in line with the training objectives and training content. Kraiger's (2002) taxonomy of cognitive, affective, and behavioral outcomes was followed as a model to increase comprehensiveness and multidimensionality of learning in the assessment. Cognitive outcomes included declarative and self-knowledge, while affective outcomes included motivation and attitude. The behavioral outcome measured involved skill acquisition, or the individual's use of learned fatigue countermeasures. In addition, we also measured reported outcomes such as fatigue, perceived sleepiness, the experience of physical symptoms, and work-family conflict.

The evaluation approach centered around a pretest-posttest design with the addition of a six-week follow-up test/survey. Methods such as IRS and RGD were used for the cognitive measures to rule out threats to validity that typically plague pretest—posttest designs and to increase confidence that trainee changes were the result of the training.

Cognitive. Declarative knowledge was assessed via recognition and recall of basic fatigue knowledge regarding causes, consequences, fatigue mitigating strategies, and appropriate situations for their use. Training-irrelevant items included information similar in nature to the training-relevant items, with a focus on a related, though different topic, but this information was not covered in training. The purpose of the declarative knowledge measure was to determine whether trainees learned the information necessary to apply fatigue countermeasures on the job and at home. Knowledge was assessed via self-report.

Affective. Attitudes regarding fatigue management were assessed via self-report, as were motivation and self-efficacy. The purpose of the attitude measure was to ensure that trainees valued fatigue management, while the motivation measure assessed whether trainees saw a need to apply fatigue management strategies. Finally, the self-efficacy measure was included to determine the extent to which trainees felt that they were capable of utilizing fatigue countermeasures.

Behavioral. The use of fatigue management strategies was assessed via open-ended, self-report and a behavioral checklist. The purpose of these measures was to determine whether trainees actually applied the training content to their daily lives.

Additional outcomes. Although not grounded in Kraiger's (2002) taxonomy of training outcomes, several other measures - including fatigue, sleepiness, the experience of physical symptoms, and work-family conflict - were also measured to determine the impact that fatigue training had on these outcomes. Fatigue was measured using a brief

self-assessment questionnaire called the Fatigue Assessment Scale (Michielsen, De Vries, & Van Heck, 2003), and sleepiness was assessed using the Epworth Sleepiness Scale, a validated tool for measuring daytime sleepiness. The physical symptoms experienced by our participants were measured using a checklist of common symptoms reportedly experienced by shiftworkers (Spector, 1987), and finally, work-family conflict was assessed using a scale that combines questions regarding how work interferes with family and how family interferes with work.

RESULTS

Due to the relatively small sample that participated in all three training evaluation phases, ANOVAs were used to identify whether significant differences existed between groups on any of the pretest measures. Results indicated that there were no significant differences between groups on pretest measures, so all three training groups were combined for further analyses.

Hypothesis Tests

Change in performance on cognitive measures following training was assessed using repeated measures ANOVAs. The overall ANOVAs were significant for acquiring new information [$F(2, 34)=70.27, p<.001$], articulating awareness [$F(2, 34)=103.83, p<.001$], and propositional knowledge [$F(1, 36, 23.11)=16.58, p<.001$]. Note that the assumption of sphericity was violated in the test of propositional knowledge, and as a result, the Greenhouse-Geisser correction is reported.

The significant overall ANOVAs were followed up by paired sample t-tests to examine the changes from pretest to posttest and from pretest to follow-up. The Bonferroni procedure was used to adjust the significance level to $p=.025$ and correct for Type 1 error. The results of these analyses are displayed in Table 1. Training produced significant gains from pretest to posttest and from pretest to follow-up across knowledge measures, thus fully supporting H1. As a result of the training, participants were better able to recognize, paraphrase, and differentiate information relevant to effective fatigue management. This effect was significant immediately following training and four to six weeks later during the follow-up evaluation.

Changes in motivation, attitude strength, and self-efficacy following training were examined using repeated-measures ANOVAs. The overall ANOVA for motivation was not significant [$F(2, 34)=2.20, p=.13, \text{partial } \eta^2=11$]. The overall ANOVA for attitude strength violated the assumption of sphericity, and as a result, the Greenhouse-Geisser correction was utilized [$F(1.44, 24.42)=3.51, p=.06, \text{partial } \eta^2=17$]. Additionally, the overall ANOVA for self-efficacy was statistically significant [$F(2, 34)=3.76, p=.03, \text{partial } \eta^2=18$].

Significant overall ANOVAs were followed up by paired sample t-tests to examine the changes from pretest to posttest and from pretest to follow-up. Note that the p value for overall test of attitude strength rounded up to .06; therefore, the decision was made to conduct paired sample t-tests for this outcome. The Bonferroni procedure was used to adjust the significance level to $p=.025$ and correct for Type 1 error. The results of these analyses are illustrated in Table 2. With the exception of motivation from pretest

Table 1.
Means, SDs, and Paired Sample t-tests of Cognitive Training Outcomes

Variable	Pretest		Posttest		Follow-up		Pre-post	Pre-follow-up
	M	SD	M	SD	M	SD	t	t
Acquiring new information	45.68	21.27	91.36	9.57	87.45	12.75	8.90*	10.08*
Articulating awareness	24.60	20.68	88.10	10.10	84.09	17.02	15.12*	10.08*
Propositional knowledge	74.07	18.28	90.37	8.32	90.00	8.63	4.00*	4.90*

Note. * $p<.001$

Table 2.
Means, SDs, and Paired Sample t-tests of Affective Training Outcomes

Variable	Pretest		Posttest		Follow-up		Pre-post	Pre-follow-up
	M	SD	M	SD	M	SD	t	t
Motivation	17.61	3.09	17.39	2.70	18.28	3.12	–	–
Attitude strength	18.78	2.34	20.06	2.31	19.94	2.69	2.64*	1.72
Self-efficacy	13.89	2.19	14.94	2.82	14.56	2.91	2.49*	1.80

Note. * $p<.025$

to posttest, all affective measures changed in the expected direction following training. The change in attitude strength and self-efficacy from pretest to posttest was statistically significant, indicating that participants felt more strongly about fatigue management and their ability to apply fatigue management strategies after participating in the training. Although attitude strength and self-efficacy showed positive effects through the follow-up period, the changes were not significant; thus, H2 is partially supported.

The application of fatigue countermeasures was assessed using a paired sample t-test. Using a checklist response format, there was a significant difference between countermeasure utilization when assessed during the pretest ($M=140.81$, $SD=13.19$) and during follow-up ($M=151.07$, $SD=13.29$), $t(17) = -3.01$, $p<.01$. Also in support of H3, prior to training 44.4% of respondents reported making changes at home, compared to 83.3% following the training. Results were similar when participants were asked about the use of fatigue countermeasures on the job, with 50% reported making changes prior to training, and 83.3 % reported making changes following the training. When asked in an open-response format, the number of strategies being used at home increased by 138.5% following training. The number of fatigue countermeasures used at work increased by 175% from pretest to follow-up.

Table 3 provides the means and standard deviations for measures of fatigue, sleepiness, physical symptoms, work-family conflict, and family-work conflict. Paired sample t-tests were conducted to determine if significant differences exist between outcomes, as measured during the pretest and follow-up. Only the Fatigue Assessment Scale demonstrated significant differences indicating that flight attendants experienced less fatigue at the time of follow-up. None of the other aforementioned outcomes were significant, so H4 is only partially supported.

Table 4 presents the means and standard deviations for the relevant and irrelevant items of acquiring new information and propositional knowledge. A 2x2 repeated measures ANOVA was used for each outcome to test the main effect of time (pretest or posttest) and relevance (relevant or irrelevant to training), as well the interaction between time and relevance.

The results of the analyses for acquiring new information indicate a significant main effect for the time factor [$F(1, 17)=33.03$, partial $\eta^2 = .66$, $p<.001$] and the relevance factor [$F(1, 17)=99.15$, partial $\eta^2 = .85$, $p<.001$]. The interaction between the time and relevance factors was also significant [$F(1, 17)=137.18$, partial $\eta^2 = .89$, $p<.001$]. These results directly support H5 by demonstrating that the difference in acquisition of new knowledge from pretest to posttest was greater for

Table 3.
Means, SDs, and Paired Sample t-tests of Additional Training Outcomes

Variable	Pretest		Follow-up		<i>t</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Federal Air Surgeon	2.56	.52	2.33	.53	1.91*
EpworthSS	9.17	2.90	8.60	3.83	0.44
Physical symptoms	36.89	9.99	35.86	9.97	0.92
Work-family conflict	14.72	3.85	15.67	4.34	-0.91
Family-work conflict	8.44	2.68	8.11	3.23	-0.47

Note. * $p<.05$, one-tailed

Table 4.
Means and SDs of Training Outcomes Based on the Internal Referencing Strategy

Variable	Relevant items				Irrelevant items			
	Pretest		Posttest		Pretest		Posttest	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Acquiring new information	45.68	21.27	91.36	9.57	33.33	18.18	37.37	16.47
Propositional knowledge	74.07	18.28	90.37	8.32	56.67	21.96	85.56	13.38

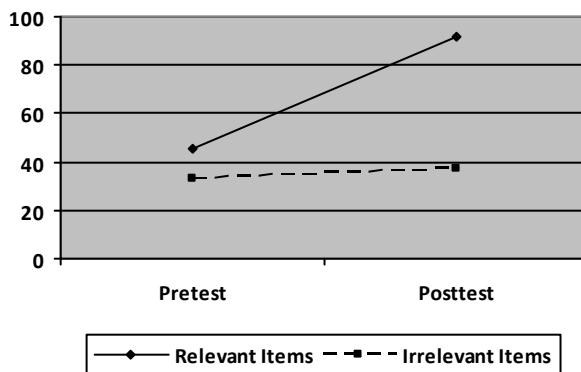


Figure 1. Change in acquiring new information by time and item relevance.

relevant items than for irrelevant items. This relationship is further illustrated in Figure 1.

The results for propositional knowledge indicate a significant main effect for the time factor [$F(1, 17)=32.91$, partial $\eta^2 = .66$, $p<.001$] and the relevance factor [$F(1, 17)=11.26$, partial $\eta^2 = .40$, $p=.004$]. The interaction between the time and relevance factors was not significant however [$F(1, 17)=2.38$, partial $\eta^2 = .14$, $p=.11$]. These results demonstrate that the difference in propositional knowledge from pretest to posttest was greater for irrelevant items than for relevant items, which does not support H5. This relationship is further illustrated in Figure 2. Possible explanations for this finding are discussed further.

The means, standard deviations, and t-tests for the RGD are presented in Table 5. To test H6, H7, and H8, the groups from training sessions 2 and 3 were compared to examine differences between pretest and posttest cognitive measures. Training session selection was based solely on the number of participants in each

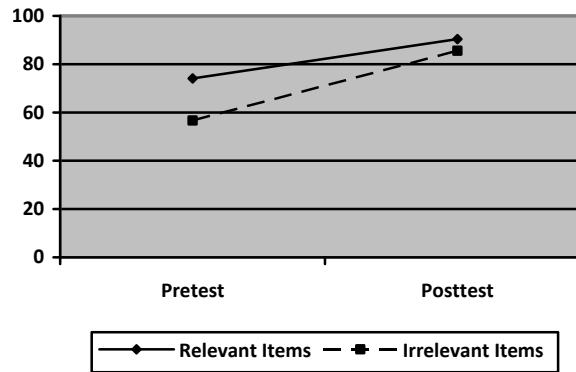


Figure 2. Change in propositional knowledge by time and item relevance.

session; sessions 2 and 3 allowed the greatest sample sizes. For the training group, performance on each cognitive outcome was examined using paired sample t-tests. All three cognitive outcomes were significant, indicating changes in knowledge between the pretest and posttest for flight attendants who participated in the training. To simulate a control group, the pretest for one of the training sessions was used as a comparison for the posttest for a training group. Differences in the cognitive measures were assessed via independent sample t-tests. As illustrated by Table 5, comparisons for all three cognitive outcomes were significant. This demonstrates significant knowledge differences between the "control" group and the post-training group. Finally, there were also significant differences between the pretest and posttest for the "control" group for all three cognitive measures. Paired sample t-tests were used to assess these differences. All analyses fully supported H6, H7, and H8 indicating that flight attendants were more knowledgeable about fatigue management as a result of training.

Table 5.

Means, SDs, and t-tests of Training Outcomes Based on the Rolling Group Design

Variable	Treatment Group				Control Group				t^a	t^b	t^c			
	Pretest		Posttest		Pretest		Posttest							
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>						
Acquiring new information	9.00	3.57	16.56	1.13	6.92	3.43	15.86	1.83	6.90*	5.33*	8.26*			
Articulating awareness	2.11	1.36	5.89	0.60	1.57	1.65	6.29	0.83	8.13*	9.19*	10.77*			
Propositional knowledge	9.22	1.30	11.89	1.45	11.14	2.35	12.79	2.04	3.77*	4.64*	2.98*			

Note. ^a Compares training pretest with training posttest. ^b Compared control pretest with training posttest.

^c Compares control pretest with control posttest. * $p < .01$, two-tailed. Training Group $n=9$ Control Group $n=14$

DISCUSSION

Overall, the results of this study demonstrate the effectiveness of a thoroughly developed and comprehensive fatigue countermeasures training program. By utilizing alternative learning outcomes and multiple evaluation strategies, we are able to gain a better understanding of the learning process and produce convergent evidence of training effectiveness. As a result of the training, participants improved their knowledge of basic fatigue information and strategy use; they acquired new information, were able to articulate awareness, and exhibited greater recognition of effective fatigue countermeasure strategies. Participants also showed improvements in their self-efficacy for addressing fatigue and the strength of their attitudes toward fatigue and the importance they place on fatigue management. In addition, and perhaps most tellingly, training participants demonstrated changes in the level of fatigue experienced and the number of fatigue countermeasure strategies they used. For example, 41.2% of flight attendants utilized naps for fatigue management following training, as compared to only 27.8% prior to training. Flight attendants even received more nightly sleep, as a result of training, increasing from 6.78 hours per night to 7.37 hours. Together, these results provide strong evidence for the effectiveness of the fatigue countermeasures training program.

Use of Kraiger, Ford, and Salas's (1993) classification of learning outcomes for the present project provided a more comprehensive understanding of the learning taking place as a result of training. Results clearly demonstrated training effectiveness in terms of cognitive learning outcomes and skill acquisition. Evaluation of affective outcomes revealed that self-efficacy and attitude strength were significantly improved following training, but that motivation was only slightly higher post-training. This finding is interesting, considering that their attitudes regarding the need to fight fatigue, and the belief that they could effectively fight fatigue, increased as a result of training. The lack of significant improvement in motivation may suggest that the information presented during training was somehow overwhelming for participants. For example, they left training feeling that fatigue was an important issue and that they were capable of applying fatigue countermeasure strategies, but perhaps the magnitude of the changes that would need to be made were simply overwhelming. Alternatively, given that there was an increase in motivation at the time of follow-up, perhaps the power for this test was lacking and the more subtle effect was undetectable.

Additional training outcomes regarding sleepiness, physical symptoms, work-family conflict, and family-work conflict were not found to be significantly different

following training. It is possible that fatigue simply does not affect these outcomes; alternatively, the four- to six-week time frame may have been insufficient to observe significant changes. This may highlight the challenges of fatigue management faced in flight operations and warrants further attention.

The present study also supports the use of alternative training evaluation strategies, including IRS and RGD. Rather than relying solely on pretest—posttest designs, which are vulnerable to the effects of history, testing, and maturation, IRS and RGD methodologies were employed in our study to provide greater confidence in the validity of the training results. Previously, IRS had only been applied to propositional knowledge or recognition of declarative knowledge on multiple choice tests (Cigularov et al., 2008; Haccoun et al., 1994). This research examined whether IRS was effective for higher level learning outcomes such as the acquisition of new knowledge. Specifically, IRS provided evidence of greater knowledge acquisition for information covered during the course of training, as opposed to information that was not a part of the training. This suggested that IRS was effective for higher level learning outcomes and can be employed more broadly as an evaluation strategy. The IRS results for propositional knowledge were not as supportive, with training-irrelevant items demonstrating improvement along with training-relevant items. In retrospect, it is likely that the topic chosen for the irrelevant items was partially to blame for improvement from pretest to posttest. Many of the same coping strategies could be applied to either topic, so when presented with multiple choice items participants were more likely to guess correctly, even though information specific to the irrelevant items had not been included in the training. Additionally, there may have been a bit of a ceiling effect for the relevant propositional items. Nearly 75% of the items were answered correctly during the pretest, and 90% were answered correctly during the posttest. It is possible that these scores did not leave enough room for improvement, thereby mitigating the effect. Overall, results of the IRS supported further use of this evaluation strategy as a method of strengthening traditional pretest—posttest designs.

The RGD also appeared to be a viable alternative for strengthening traditional training evaluation designs. As hypothesized, results indicated meaningful differences between pretest measures of a designated control group and the posttest measures of a training group. Use of a control group that eventually completed training allowed us to have greater confidence in the training results and helped to protect against potential threats such as testing effects, history, or maturation. While this evaluation design is not widely cited in the empirical literature, it

certainly poses an alternative for real-world settings where constraints inhibit the use of actual control groups or other, more thorough evaluation designs.

Given the multi-industry development of this training program, it is likely that the positive effects will generalize to other populations that deal with similar non-traditional schedules and other occupational conditions that contribute to fatigue. Although tailored toward the specific challenges faced by flight attendants, much of the training information represents basic knowledge about fatigue and how to effectively prevent and manage it. It seems highly likely that this training program would be useful and effective across many industries. Given the effects fatigue may have on safety-related behavior and the potential for workplace incidents or accidents, fatigue countermeasures training should be an effective prevention strategy for many organizations (Caldwell, 2005; Rosekind et al., 1996). Only 14.6% of flight attendants reported having received any fatigue education or training, but nearly all reported that they experienced fatigue. The results from this comprehensive training program, as well as others, suggest that it is an effective strategy for reducing fatigue and promoting other positive outcomes. Taken together, this suggests that fatigue countermeasures training should be utilized more frequently as an intervention strategy for employees with non-traditional schedules.

Future research should consider the use of fatigue countermeasure training across modes of operations. Additionally, the benefits of training should be examined in the context of improvements in safety-related behavior and achievement of organizational objectives. This was not possible in the current study, but it has implications for the widespread use of the training. These training materials should be considered for application via computer-based training to enhance usability and cost-effectiveness. Finally, researchers might explore the use of IRS with skill acquisition or behavioral outcomes. To date, there is no existing research examining the suitability of IRS for determining behavioral outcomes.

The computer-based fatigue countermeasures workshop for flight attendants is available at http://lfclients2.com/Clients/FAA/FatigueFA/Published/EXE/FAA_Fatigue_FA.zip.

REFERENCES

Aguinis, H. & Branstetter, S.A. (2007). Teaching the concept of the sampling distribution of the mean. *Journal of Management Education, 31*, 467-483.

Avers, K., Hauck, E., Blackwell, L., & Nesthus, T. (2009a). Flight attendant fatigue, part VI: Fatigue countermeasure training and potential benefits. (Report No. DOT/FAA/AM-09/20). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Avers, K., King, S., Nesthus, T., Thomas, S., & Banks, J. (2009b). Flight attendant fatigue, part I: National duty, rest, and fatigue survey. (Report No. DOT/FAA/AM-09/24). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Banks, J., Avers, K., Nesthus, T., & Hauck, E. (2009). Flight attendant fatigue, part V: A comparative study of international flight attendant fatigue regulations and collective bargaining agreements. (Report No. DOT/FAA/AM-09/22). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Caldwell, J.A. (2005). Fatigue in aviation. *Travel Medicine and Infectious Disease, 3*, 85-96.

Cigularov, K., Chen, P., Thurber, B.W., & Stallones, L. (2008). Investigation of the effectiveness of school-based suicide education program using three methodological approaches. *Psychological Services, 5*, 262-274.

Frese, M., Beimel, S., & Schoenborn, S. (2003). Action training for charismatic leadership: Two evaluations of studies of a commercial training module on inspirational communication of a vision. *Personnel Psychology, 56*, 671-697.

Haccoun, R.R., & Hamiaux, T. (1994). Optimizing knowledge tests for inferring acquisition levels in single group training evaluation designs: The internal referencing strategy. *Personnel Psychology, 47*, 593-604.

Holcomb, K., Avers, K., Dobbins, L., Banks, J., Blackwell, L., & Nesthus, T. (2009). Flight attendant fatigue, part IV: Analysis of incident reports. (Report No. DOT/FAA/AM-09/25). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Kraiger, K. (2002). Decision-based evaluation. In Kraiger, K. (Ed), *Creating, implementing, and managing effective training and development* (331-375). San Francisco: Jossey-Bass.

Kraiger, K., Ford, J.K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of Applied Psychology*, 78, 311-328.

Michielsen, H.J., De Vries, J., & Van Heck J.L. (2003). Psychometric qualities of a brief self-rated fatigue measure: The Fatigue Assessment Scale. *Journal of Psychosomatic Research*, 54, 345-352.

Nesthus, T.E., Schroeder, D.J., Connors, M.M., Rentmeister-Bryant, H.K., & DeRoshia, C.A. (July, 2007). Flight attendant fatigue. (Report No. DOT/FAA/AAM- 07/21). Washington, DC: Office of Aerospace Medicine, Federal Aviation Administration.

Oostrom, J.K., & van Mierlo, H. (2008). An evaluation of an aggression management training program to cope with workplace violence in the healthcare sector. *Research in Nursing & Health*, 31, 320-328.

Quinones, M.A., & Tonidandel, S. (2003). Conducting training evaluation. In J.E. Edwards, J.C. Scott, & N.S. Raju (Eds.) *The human resources program-evaluation handbook* (225-243). London: Sage.

Roma, P., Mallis, M., Hursh, S., Mead, A., & Nesthus, T. (2010). Flight attendant fatigue recommendation II: Flight attendant work/rest patterns, alertness, and performance assessment. (Report No. DOT/FAA/AM-10/22). Washington, DC: Federal Aviation Administration, Office of Aerospace Medicine.

Rosekind, M.R., Gander, P.H., Gregory, K. B., Smith, R.M., Miller, D.L., Oyung, R., Webbon, L.L., & Johnson, J.M. (1996). Managing fatigue in operational setting 1: Physiological consideration and countermeasures. *Behavioral Medicine*, 21, 157-165.

Shadish, W.R., Cook, T.D. & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin Company.

Spector, P.E. (1987). Interactive effects of perceived control and job stressors on affective reactions and health outcomes for the clerical workers. *Work and Stress*, 1, 155-162.

